Arch. Min. Sci., Vol. 58 (2013), No 4, p. 1317–1331

Electronic version (in color) of this paper is available: http://mining.archives.pl

DOI 10.2478/amsc-2013-0091

TOMASZ NIEDOBA*, DARIUSZ JAMRÓZ**

VISUALIZATION OF MULTIDIMENSIONAL DATA IN PURPOSE OF QUALITATIVE CLASSIFICATION OF VARIOUS TYPES OF COAL

WIZUALIZACJA WIELOWYMIAROWYCH DANYCH W CELU KLASYFIKACJI JAKOŚCIOWEJ RÓŻNYCH TYPÓW WĘGLA

Coal as energetic raw material features by many parameters determining its quality. In classification of coal types there are many of them with typical division of energetic, semi-coking and coking coal. The data concerning coal are usually treated as independent values while this kind of approach is not always right. Authors proposed new solutions in this aspect and performed the multidimensional analysis of three selected types of coal featuring by various properties which originated from three various hard coal mines located in Upper Silesia Region. The object of the research was so-called raw coal which was not processed before. For each type of coal the detailed statistical analysis of seven chosen properties of coal was performed. To perform adequate and complete statistical analysis it is necessary to analyze the chosen properties of coal together in multidimensional way. It was decided to apply new and modern visualizing methods of multidimensional data which were observational tunnels method and parallel coordinates method. The applied methods allowed to obtain visualization of seven-dimensional data describing coal. By means of these visualizations it was possible to observe the significant division of the features space between researched types of coal. These methods allowed to look at the investigated data from various perspectives and make possible to determine significant differences between researched materials. For the investigated coals such differences were determined clearly what proved that by means of these methods it is possible to successfully identify type of coal as well to analyze in details its individual properties and identify, for example, particle size fraction etc. The obtained results are innovative and are the basis for more detailed researches taking into consideration also other coal properties, including its structure and texture. This methodology can be also applied successfully for other types of raw materials, like ores.

Keywords: multidimensional analysis, observational tunnels, parallel coordinates, coal, mineral processing, coal energy

* AGH UNIVERSITY OF SCIENCE AND TECHNOLOGY, FACULTY OF MINING AND GEOENGINEERING, DEPARTMENT OF ENVIRONMENTAL ENGINEERING AND MINERAL PROCESSING, AL. A. MICKIEWICZA 30, 30-059 KRAKOW, POLAND, E-mail: tniedoba@agh.edu.pl

** AGH UNIVERSITY OF SCIENCE AND TECHNOLOGY, FACULTY OF ELECTRICAL ENGINEERING, AUTOMATICS, COMPUTER SCIENCE AND BIOMEDICAL ENGINEERING, DEPARTMENT OF APPLIED COMPUTER SCIENCE, AL. A. MICKIEWICZA 30, 30-059 KRAKOW, POLAND, E-mail: jamroz@agh.edu.pl

Surowce mineralne, które podlegają wzbogacaniu w celu ich lepszego wykorzystania mogą być (charakteryzują się) charakteryzowane wieloma wskaźnikami opisującymi ich, interesujące przeróbkarza, cechy. Podstawowymi cechami są wielkość ziaren oraz ich gęstość, które decydują o przebiegu rozdziału zbiorów ziaren (nadaw) i efektach takiego rozdziału. Rozdział prowadzi się z reguły, w celu uzyskania produktów o zróżnicowanych wartościach średnich wybranej cechy, która zwykle charakteryzowana jest zawartością określonego składnika surowca wyznaczoną na drodze analiz chemicznych. Takie podejście do surowca mineralnego prowadzi do potraktowania go jako wielowymiarowego wektora $X = [X_1, \ldots, X_n]$. Zasadniczym problemem jest także wybór jednostki populacji generalnej (ziarno, jednostka objętości lub masy), co może decydować o kierunkach charakteryzowania wielowymiarowych powiązań cech wektora X. Takimi kierunkami charakteryzowania mogą być:

- wielowymiarowe rozkłady wektora losowego X wraz ze wszystkimi konsekwencjami metody (Lyman, 1993; Niedoba, 2009; 2011; Olejnik et al., 2010; Niedoba & Surowiak, 2012);
- wielowymiarowe równania regresji wraz z analizą macierzy współczynników korelacji liniowej oraz korelacji cząstkowej (Niedoba, 2013);
- analiza czynnikowa (Tumidajski & Saramak, 2009);
- inne metody, w tym wizualizacja metodą tuneli obserwacyjnych (Jamróz, 2001), osi równoległych oraz wizualizacja zależności pomiędzy wielowymiarowymi bryłami (Jamróz, 2009).

Wielowymiarowe rozkłady wektora X traktowanego jako wektor losowy, mają już swoją bogatą literaturę i praktyczne ich zastosowanie i nie będą przedmiotem tej publikacji. Pozostałe metody są ze sobą w pewien sposób powiązane, co skrótowo zostało przedstawione w artykule.

Macierze współczynników korelacji liniowej i współczynników korelacji cząstkowej są związane, z reguły, z istniejącymi modelami liniowymi zależności występujących między badanymi zmiennymi wektora X. Współczynniki korelacji liniowej są wyznaczane dla par zmiennych losowych całkowicie niezależnie od pozostałych zmiennych. Cząstkowe współczynniki korelacji liniowej wyznaczane są w oparciu o macierz współczynniki korelacji liniowej z uwzględnieniem roli pozostałych zmiennych w rozważanym równaniu regresji liniowej. W przypadku analizy trzech zmiennych losowych, z których jedna jest traktowana jako zmienna zależna a dwie pozostałe jako niezależne sprowadza się to do wyznaczania współczynników korelacji dla zrzutowanych punktów równolegle do płaszczyzny regresji na ściany układu współrzędnych. Pozwala to wyznaczyć hierarchię (siłę wpływu) zależności zmiennych w rozpatrywanym układzie. Na analizie macierzy współczynników korelacji liniowej oparta jest analiza czynnikowa, która pozwala pogrupować występujące zmienne w tzw. czynniki, które reprezentują połączone wpływy zmiennych na rezultaty rozpatrywanych procesów, czyli przeprowadzić pewną klasyfikację zmiennych.

W klasyfikacji typów węgli wyróżnia się wiele typów, z umownym podziałem na węgle energetyczne i koksujące. Dane dotyczące węgla są traktowane zwykle jako niezależne wielkości, przy czym takie podejście nie zawsze jest właściwe. Autorzy zaproponowali nowe rozwiązania w tym zakresie i dokonali wielowymiarowej analizy trzech wybranych typów węgla o różnych właściwościach (węgle typu 31, 34.2 oraz 35), które pochodziły z trzech różnych kopalń zlokalizowanych w Górnośląskim Okręgu Przemysłowym. Obiektem badań w każdej z tych kopalń był tzw. węgiel surowy, nie poddawany procesom przeróbczym. Dla każdego z węgli dokonano szczegółowej analizy wybranych siedmiu cech, opisujących jego właściwości, których przykładowe wyniki zostały zaprezentowane w tabelach 1-3. Aby dokonać adekwatnej i dokładnej analizy statystycznej zebranych danych konieczna jest wielowymiarowa analiza wybranych cech węgla łącznie. Zdecydowano się na zastosowanie nowatorskich metod wizualizacji wielowymiarowych danych, którymi były metoda tuneli obserwacyjnych oraz metoda osi równoległych. Zasady i metodyka badań zostały przedstawione w podrozdziałach 2 i 3. Zastosowane metody umożliwiły uzyskanie wizualizacji siedmiowymiarowych danych opisujących węgiel. Za pomocą tych wizualizacji możliwe jest zaobserwowanie wyraźnego podziału przestrzeni cech pomiędzy badanymi typami węgla. Metody te umożliwiły spojrzenie na badane dane z różnych perspektyw, które pozwalają na stwierdzenie zasadniczych różnic badanych materiałów. Dla badanych węgli stwierdzono wyraźne takie różnice co świadczy o tym, że za pomocą proponowanych metod możliwa jest skuteczna identyfikacja typu węgla, jak również dokładniejsza analiza jego poszczególnych cech i identyfikacja np. klasy ziarnowej. Szczegółowe obrazy i ich interpretacja zostały przedstawione w rozdziale 3 i we wnioskach końcowych. Rysunki 3-5 obrazują różnice pomiędzy poszczególnymi typami węgla otrzymane metodą tuneli obserwacyjnych. Wyraźnie można rozgraniczyć próbki dotyczące poszczególnych węgli a tym samym możliwa jest identyfikacja typu węgla na podstawie wielowymiarowej analizy. Rysunki 6-7 pokazują zastosowanie innej metody wielowymiarowej, którą była metoda osi równoległych. Metoda ta okazała się być skuteczna do uzyskania informacji o konieczności przeskalowania poszczególnych cech, w celu uzyskania bardziej

czytelnych rezultatów. Natomiast rysunek 10 pokazuje różnice otrzymane metodą tuneli obserwacyjnych pomiędzy charakterystykami konkretnych klas ziarnowych wybranego materiału, którym w tym przypadku był węgiel typu 31. Uzyskane wyniki i zastosowana metodyka są nowatorskie i stanowią bazę pod bardziej szczegółowe badania, biorące pod uwagę także inne charakterystyki węgli, w tym ich strukturę i teksturę. Za pomocą przedstawionych metod możliwe jest stwierdzenie, czy wybrane cechy są wystarczające do identyfikacji zarówno typu węgla, jak również klasy ziarnowej i innych jego cech. Metodyka ta może być również stosowana z powodzeniem dla innych typów surowców mineralnych, np. dla rud.

Słowa kluczowe: analiza wielowymiarowa, tunele obserwacyjne, osie równoległe, węgiel, przeróbka surowców mineralnych, energia z węgla

1. Introduction

Mineral raw materials which are beneficiated in purpose of their better utilization can be characterized by multiple factors describing their various features. The basic ones are particles size and their density which decide about the course of separation of particles sets (feeds) and effects of such process (Brożek & Surowiak, 2010). The separation process is conducted usually in purpose of obtaining products of various mean values of selected features which is normally characterized by contents of selected component determined by chemical analyzes. Such approach to mineral raw material leads to treat it as multidimensional random vector $X = [X_1, \ldots, X_n]$. The main problem is also to select individual units of general population (single particle, unit of volume or mass etc.) what can decide about the ways of characterizing multidimensional connections between features of vector X. Such ways are:

- Multidimensional distribution functions of random vector X with all consequences of this method (Lyman, 1993; Niedoba, 2009; 2011; Olejnik et al., 2010; Niedoba & Surowiak, 2012);
- Multidimensional regressive equations with analysis of matrix of coefficients of linear correlation and partial correlation (Niedoba, 2013; Tumidajski & Saramak, 2009);
- Factor analysis (Stanisz, 2007;Tumidajski & Saramak, 2009);
- Other methods including visualization by observational tunnels method (Jamróz, 2001), parallel coordinates and visualization of relations between multidimensional blocks (Jamróz, 2009).

The multidimensional distributions of vector X treated as random vector and their practical applications are widely described in the literature and will not be the object of this paper. The other methods are connected somehow what is presented in this paper.

The matrices of coefficients of linear and partial correlations are usually connected with existing linear models of relations between researched random variables of vector X. The coefficients of linear correlation are determined for pairs of random variables totally undependably on other variables. The partial correlation coefficients are determined on the basis of the matrix of coefficients of linear correlation with taking into consideration role of other variables in certain equation of linear regression. In case of analysis of three random variables from which one is treated as dependent variable and two others as independent ones it leads to determination of correlation coefficients for projections of points being parallel to regressive plane. It allows to determine the hierarchy (power of influence) of relations between variables in researched system. On the basis of matrix of linear coefficients of correlation the factor analysis can be performed which allow to group the existing variables into so-called factors representing joined influences of variables on the results of investigated processes. So, some sort of classification is to conduct.

In the presented paper the methods of visualization of multidimensional data are presented which also allow to conduct the comparisons of the researched data sets and to determine possibilities of their classification. They are then some sort of continuation and development of the methods discussed above.

2. General principles of visualization of multidimensional data

The qualitative analysis of multidimensional data (properties of material) obtained by empirical experiments results can be given by application of multidimensional visualization method. The results of these analyzes can be helpful by materials characteristics as well by construction of mineral processing models based on this data.

Attempts to depict multidimensional data have been undertaken on many occasions. One of the methods applied to provide a visual image of multidimensional data was the grand-tour method, which created a continuous 1-parameter family of d-dimensional projections of n-dimensional data. It was described by Asimov (Asimov, 1985), for the first time. Then, it was developed by many authors (Buja & Asimov, 1985; Cook et al., 1995; Hurley & Buja, 1990). The method of principal component analysis (Li et al., 2000) makes use of an orthogonal projection of the observation set into a plane represented by specially chosen vectors which are the eigenvectors corresponding with the two highest eigenvalues of the covariance matrix of the observation set. The use of neural networks for data visualization (Jain & Mao, 1992; Kraaijveld et al., 1995; Mao & Jain, 1995; Aldrich, 1998) is based on the process of transforming n-dimensional data space into a 2-dimensional space by applying a neural network. To visualize multidimensional data, a parallel coordinates method (Chatterjee et al., 1993; Gennings et al., 1990; Inselberg, 1985; Inselberg et al., 1994; Wegman, 1990; Chou et al., 1999) was also applied. In this method parallel coordinates are placed on a given plane at a uniform rate. Similar to this is the method of multidimensional data visualization which uses a star graph (Sobol & Klein, 1989). In this method, n coordinates radiate from a central point, dividing a circle into n equal parts. Multidimensional scaling (Kim et al., 2000) is also used to visualize multidimensional data where points are transferred to a space with a lower number of dimensions so as to make the distance between any two points as similar as possible to their previous proximity. In the scatter-plot matrices method (Becker et al., 1987; Cleveland, 1984; Eick & Wills, 1995), a set of multidimensional data is presented with a series of 2-dimensional dependencies. In a method using the so-called *relevance maps* (Assa et al., 1997; 1999) special points F_1, F_2, ..., F_n, representing individual features are placed on the data visualization plane. The layout of the points representing the multidimensional data are presented reflects the relations between these data and the features so that the more significant the feature is in an object, the closer the point representing the point F_i object should be placed. Another method of visualization with regard to multidimensional data is the *mosaic plots method* (Hartigan & Kleiner, 1981; Heike, 2000) which is a natural extension of 1-dimensional bar charts. Visualization of multidimensional solids is also possible (Jamróz, 2009). The observational tunnels method (Jamróz, 2001) makes it possible to achieve an external view of the observed multidimensional sets of points using tunnel radius, introduced by the author.

2.1. Observational Tunnels Method

Theoretical grounds of Observational Tunnels method were described in paper (Jamróz, 2001). Intuitively, it may be said that the method of observational tunnels makes use of a parallel projection with a local orthogonal projection of an extent limited by the *maximal radius of the tunnel.* This solution makes it possible to observe selected parts of a space bearing important information, which is not possible using an orthogonal projection, for example. The method of projection used in this paper is presented in a demonstrative manner on Fig. 2. The *observational plane P* will be used as a screen through which any object placed in *space X* will be viewed. This ***observational plane P*** $\subset X$ is defined as: $P = \delta(w,\{p_1,p_2\})$, where:

$$\delta\left(w,\{p_1,p_2\}\right) \overset{def}{=} \{x \in X : \exists \beta_1, \beta_2 \in F,\ \textbf{such that}\ \ x = w + \beta_1 p_1 + \beta_2 p_2\} \qquad (1)$$

X is any n-dimensional ($n \geq 3$) vector space, over an F field of real numbers, with a scalar product.

Vector w will indicate the position of the screen midpoint, whereas p_1,p_2 will indicate its coordinates. Let assume for the moment, that the *space X* is 3-dimensional (an example assuming a space with more dimensions would be more difficult to conceive) and that *observational plane P* is 1-dimensional (i.e. it is possible to observe the pertinent reality not through a segment of a 2-dimensional plane but through a segment of a line). Additionally, let take a vector r, being the *proper direction of projection* onto the *observational plane P* (The ***proper direction of projection r*** onto the *observational plane* $P = \delta(w,\{p_1,p_2\})$ is defined as any vector $r \in X$ if vectors $\{p_1,p_2,r\}$ are an orthogonal system). Let's determine $k_{a,r}$ (i.e. a *line parallel* to r and passing through a) for *observed point a*. As shown on Fig. 1, the line $k_{a,r}$ need not have common points with P. However, $k_{\mathrm{a,r}}$ always has one common point with *hypersurface S* containing P and being orthogonal to r. (the ***hypersurface*** $\boldsymbol{S_{(s,d)}}$, anchored in $s \in X$ and directed towards $d \in X$ is defined as:

$$S_{(s,d)} \overset{def}{=} \left\{x \in X : (x - s, d) = 0\right\} \qquad (2)$$

A line parallel to r and passing through a does not have to have common points with P, however, it always has exactly one common point with hypersurface S containing P and being orthogonal to r. In the above mentioned case only point a_2 will be visible using observational plane P.

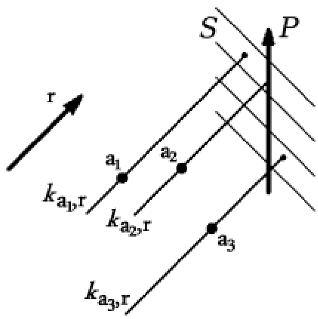

Fig. 1. Presentation of the rules of projection on plane P in observational tunnels method

In practice, only at some particular orientations of *observational plane P* could *some* points be viewed. This implies that in the majority of cases, when viewing a set of points using *observational plane P* nothing could be seen. In order to avoid such a situation, let us assume that the points visible on *observational plane P* do not only include points situated on lines parallel to r and passing through P, but also the points which are situated on lines parallel to r and passing through S (i.e. the *hypersurface* containing P and orthogonal to r) within a smaller distance from *observational plane P* than a certain fixed value. This distance for *observed point a* will be represented by vector b_a called the *tunnel radius*:

$$b_a = \psi r + a - w - \beta_1 p_1 - \beta_2 p_2, \tag{3}$$

where:

$$\psi = \frac{(w-a\,,r)}{(r\,,r)}, \qquad \beta_1 = \frac{(\psi r + a - w,\ p_1)}{(p_1\,,p_1)}, \qquad \beta_2 = \frac{(\psi r + a - w,\ p_2)}{(p_2\,,p_2)} \tag{4}$$

$r \in X$ — a *proper direction of projection* onto *observational plane P*.

In the case presented on Fig. 2, at the point *e* of *observational plane P*, all points situated in the *tunnel* whose intersection is a segment of and which is spreading along *r* will be visible. However, generally, at the point *e* on the *observational plane P*, all points situated in the *tunnel* whose intersection is an *n*-3 dimensional sphere and spreading along the *direction of projection r* will be visible.

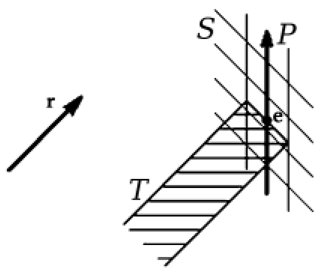

Fig. 2. Way of choosing observational tunnel *T*

Tunnel *T* for point e is shown. (The area hatched with horizontal lines). All points that belong to tunnel *T* will be visible at point e of observational plane *P*.

2.2. The Drawing Procedure

The algorithm below should be followed in order to draw the *projection of observed point a* consistent with the *direction of projection r* onto *observational plane* $P = \delta(w,\{p_1,p_2\})$:

1. the *distance of projection of observed point a* is to be calculated using the formula: $\psi = (w - a,r)/(r,r)$
2. the *position of the projection* (i.e. the pair $\beta_1,\beta_2 \in F$) *of observed point a* is to be calculated using the formula: $\beta_1=(\psi r + a - w,p_1)/(p_1,p_1)$, $\beta_2=(\psi r + a - w,p_2)/(p_2,p_2)$

3. the *tunnel radius* b_a *of point a* is to be calculated using the definition:
   $b_a = \psi r + a - w - \beta_1 p_1 - \beta_2 p_2$
4. at this point it should be verified whether the scalar product (b_a,b_a) is lower than the maximum *tunnel radius* determined at a given time and whether the *distance of the projection of observed point a* is shorter than the maximum range of view determined at a given time. If this is the case, then one should draw a point on observational plane $P = \delta(w,\{p_1,p_2\})$ in the position of coordinates (β_1,β_2), otherwise the point should not be drawn.

The scalar product is to be calculated using the formula: $(x,y) = \sum_{i=1}^{n} x_i y_i$, where: $x = (x_1,x_2,\ldots x_n)$, $y = (y_1,y_2,\ldots y_n)$, n-number of dimensions, $n \geq 3$.

3. Experiment

Three types of coal, types 31 (energetic coal), 34.2 (semi-coking coal) and 35 (coking coal) in Polish classification were used in the investigation. They originated from three various Polish coal mines and all of them were initially screened on a set of sieves of following sizes: –1.00, –3.15, –6.30, –8.00, –10.00, –12,50, –14.00, –16.00 and –20.00 mm. Then, the size fractions were additionally separated into density fractions by separation in dense media using zinc chloride aqueous solution of various densities (1.3, 1.4, 1.5, 1.6, 1.7, 1.8 and 1.9 g/cm^3). The fractions were used as a basis for further consideration and additional coal features were determined by means of chemical analysis. For each density-size fraction such parameters as combustion heat, ash contents, sulfur contents, volatile parts contents and analytical moisture were determined giving together with mass of these fractions seven various features for each coal. The examples of such data were presented in tables 1-3 showing the data for density-size fractions 6.3-3.15 mm for each type of coal.

TABLE 1

Data for density-size fraction 6.3-3.15 mm – coal, type 31

6.3-3.15 [mm]						
Density [g/cm^3]	**Mass**	**Combustion heat [cal]**	**Ash contents [%]**	**Sulfur contents [%]**	**Volatile parts contents V^a**	**Analytical moisture W_a**
<1.3	3575.6	7431	1.28	0.65	37.04	3.18
1.3-1.4	2611.4	7031	4.23	0.67	32.74	3.94
1.4-1.5	341.6	6144	14.84	1.36	27.91	3.60
1.5-1.6	128.2	5568	23.1	1.44	26.32	3.21
1.6-1.7	108	4317	37.01	1.64	24.51	2.55
1.7-1.8	104	3874	41.57	1.75	25.17	2.31
1.8-2.0	88.2	3658	42.63	1.81	24.89	2.19
>1.9	2496.8	3445	43.33	1.95	24.33	2.08

TABLE 2

Data for density-size fraction 6.3-3.15 mm – coal. type 34.2

6.3-3.15 [mm]						
Density [g/cm^3]	**Mass**	**Combustion heat [cal]**	**Ash contents [%]**	**Sulfur contents [%]**	**Volatile parts contents V^a**	**Analytical moisture W_a**
<1.3	1817.9	8355	1.13	0.31	31.03	1.11
1.3-1.4	503.6	7975	4.92	0.42	26.41	0.93
1.4-1.5	69.5	6775	16.21	0.69	25.52	0.83
1.5-1,6	46,7	5887	24.53	0.79	25.01	0.92
1,6-1,7	30,6	5001	31.01	0.67	26	0.92
1,7-1,8	21,4	4402	25.34	0.68	25.34	1.31
1,8-1,9	3,6	4428	36.86	1.11	26.41	0.92
>1,9	248,2	731	79.9	0.43	12.19	0.75

TABLE 3

Data for density-size fraction 6.3-3.15 mm – coal, type 35

6,3-3,15 [mm]						
Density [g/cm^3]	**Mass**	**Combustion heat [cal]**	**Ash contents [%]**	**Sulfur contents [%]**	**Volatile parts contents V^a**	**Analytical moisture W_a**
<1.3	2081.1	8316	2.49	0.40	21.25	1.30
1.3-1.4	745.2	7772	8.30	0.5	19.28	1.19
1.4-1.5	294.4	6647	18.98	0.59	18.62	1.14
1.5-1.6	153.7	5795	27.90	0.61	17.82	1.29
1.6-1.7	136.3	4985	34.64	0.62	17.24	1.5
1.7-1.8	92.4	4245	41.20	0.68	16.75	1.46
1.8-1.9	95.9	3584	48.57	0.79	15.48	1.65
>1.9	1898.5	964	77.16	0.39	10.86	1.45

3.1. Visualization of 7-dimensional data

The observational tunnels method was used to analyze the data describing three types of coal. This method served to visualize the multidimensional data. Each of seven parameters was treated as separated dimension. So, the 7-dimensional space was created. The figures 3-5 present the experimental results taking into consideration data of all three types of coal: 34.2 (61 samples), 35 (72 samples) and 31 (72 samples). Lower amount of samples for coal of type 34.2 occurred from the fact that some results for this coal were missing. Figure 3 presents the view from which it is visible that the data concerning coal of type 31 are located in other part of the space than coal of type 35.

On the basis of these figures it can be stated that the accepted parameters are sufficient to properly identify whether certain sample originated from coal of type 31 or 35. Figure 4 presents the view allowing to state that data concerning coal of type 35 are located in other part of the space than coal of type 34.2. From these three views it can be stated then that the accepted parameters are sufficient to correctly identify whether certain sample originated from coal of type 31, 34.2 or 35. At the same moment it was impossible to achieve one view from which such conclusion would be clear. It means that the analyzed data are complicated.

Fig. 3. Obtained view from which it occurs that data for coal 31 (gray spots) gather in other part of the space than coal 35 (bright spots – gathered in two groups/subareas of the space)

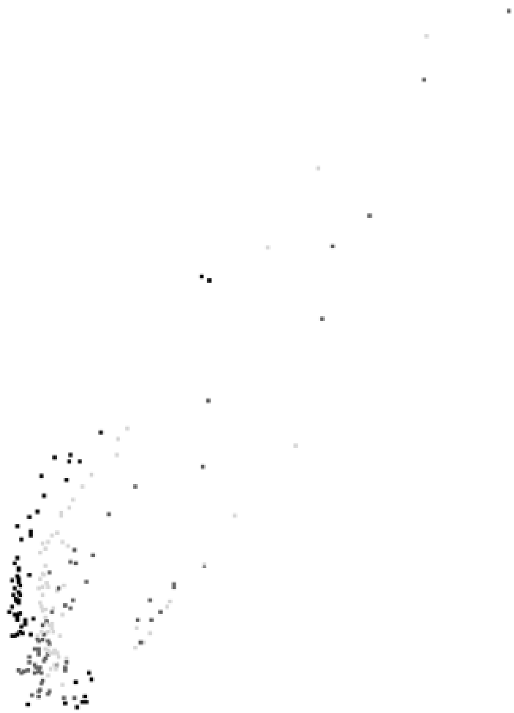

Fig. 4. View from which it occurs that data for coal 35 (bright spots) gather in other part of the space than coal 34.2 (black spots)

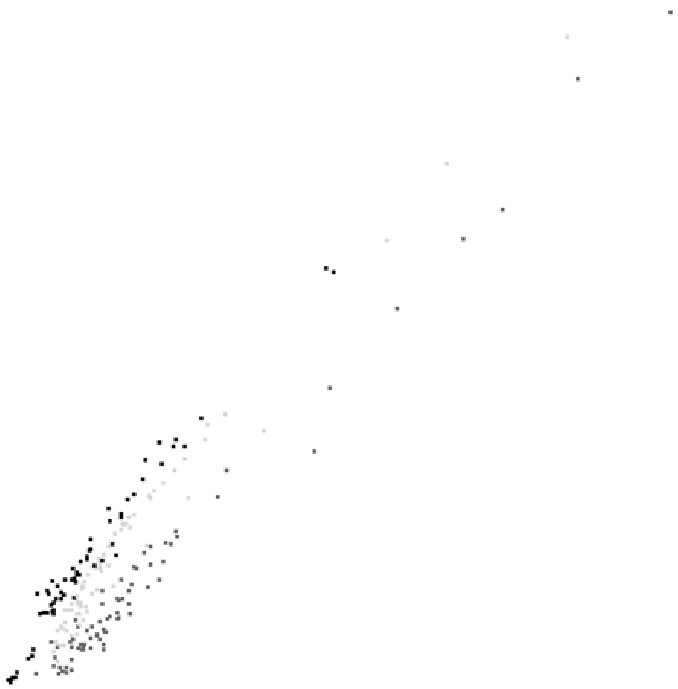

Fig. 5. View from which it occurs that data for coal 31 (Gray spots) gather in other part of the space than coal 34.2 (black spots)

Furthermore, to visualize the data presented above the method of parallel coordinates was used too. In this method there are *n* parallel axes responding to *n* dimensions of the space. One point of the space is represented by broken curve. This curve crosses through each i^{th} axis in spot which represents i^{th} coordinate of the point. Figure 6 presents the general idea of this method.

Figure 7 presents the view of the analyzed 7-dimensional set of data by means of parallel coordinates method. On the basis of this view it is impossible to state whether samples of individual coals can be separated or not. However, it can be said that before further analysis the data should be once again scaled in the way assuring change of the range of values in each coordinate of the space in the similar way. This would allow to obtain the situation where undependably on the choice of metrics each of the parameters would have similar weight by evaluation or classification of the data.

3.2. Visualization of 7-dimensional data: coal 31 with particle size fractions

The view of data representing coal of type 31 with particle size fractions is presented on figure 8. On this picture 9 fractions – points of certain brightness – represent samples of the same fraction. By such approach it was impossible to achieve views from which conclusions can be made.

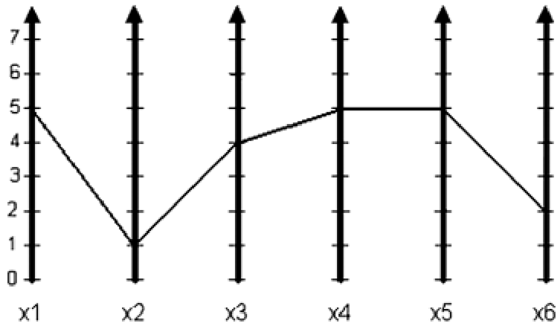

Fig. 6. Parallel coordinates method. On example, 6-dimensional point of coordinates (5, 1, 4, 5, 5, 2) is represented as broken line consisted of 5 sections joining 6 points on the surface, by one point on each parallel axis. The location of the point on i^{th} axis responds to i^{th} coordinate of 6-dimensional point

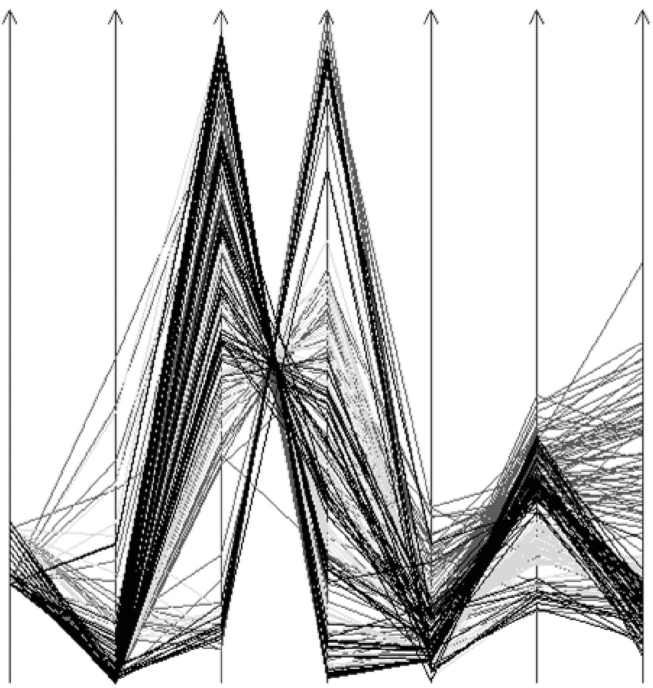

Fig. 7. View of data for coal 31 (gray spots), coal 34.2 (black spots) and coal 35 (bright spots)

The conclusions were made by approach where the separation between fractions was analyzed in pairs. On figure 9 such pair is presented where fractions 20-16 mm and 14-12.5 mm were selected. On the basis of this view it can be stated that the points representing these fractions are located in separated subareas of the space. It can be said then that the accepted parameters are sufficient to proper identification whether certain sample origins from particle size fraction

Fig. 8. View of data for coal 31 with particle size fractions. Points of various brightness represent various fractions. Too large number of fractions cause the view not clear

20-16 mm or 14-12.5 mm. On figures 10 the views for all remained pairs of each fraction and fraction 20-16 mm are presented. On the basis of each such figure it can be stated that the accepted parameters are sufficient to proper identification of particle size fraction from which the sample originated. It can be clearly stated because it is visible on figures that each fraction in each pair is located in separated subareas of the space. If then particle size fraction 20-16 mm with each other fraction is located in separated subarea of the space then it means that the accepted parameters are sufficient to properly identify whether the sample origins from particle size fraction 20-16 mm or not.

4. Conclusions

On the basis of visualization of 7-dimensional data by observational tunnels method the view for All pairs of coal types were obtained from which it occurred that the accepted parameters are sufficient for the proper identification of the coal sample origin.

One view for all three types of coals together from which conclusion from the previous point would be true was not possible to achieve. It proves that the structure of the analyzed data is complicated.

On the basis of visualization of 7-dimensional data, for individual coal types with particle size fractions it was not possible to achieve clear views for all fractions together. The conclusions were obtained by analysis of pairs of particle fractions. On the basis of achieved views it can

Fig. 9. View of data for coal 31, two particle size fractions: 20-16 mm (bright spots) and 14-12.5 mm (black spots). The obtained view allows to state that the fractions gathered in separated parts of the space

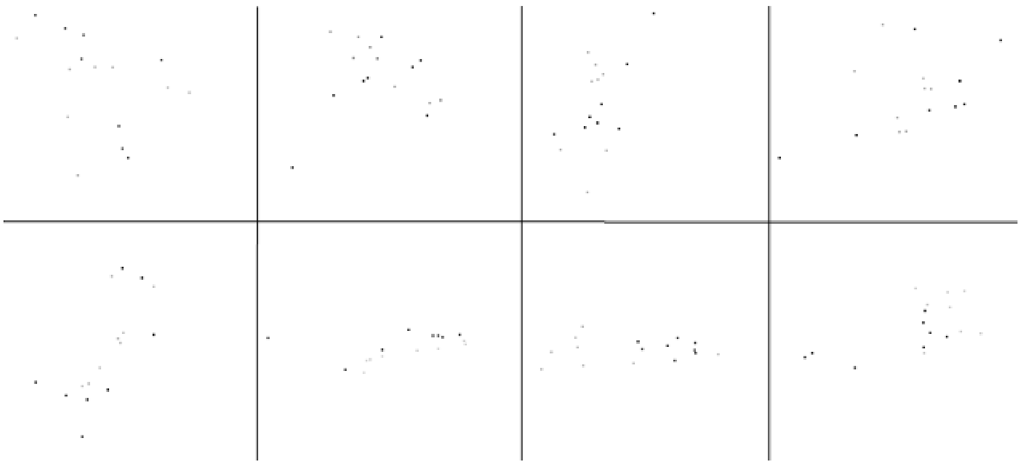

Fig. 10. The joined pictures of views for coal 31 with selected two particle fractions

be stated that the accepted parameters are sufficient for proper identification of sample particle origin considering particle size fraction.

From the view obtained by parallel coordinates method for all types of coal together it was impossible to conclude if the samples of individual coal types can be separated or not. Before the further analysis it is worthy to scale the data in the way assuring for each parameter similar weight.

The results of visualization obtained by observational tunnels method proved that on the basis of these data it is possible to obtain proper mathematical models describing mineral processing of raw materials.

ACKNOWLEDGEMENT

The paper is the effect of the scientific project no. *N N524 339040*, no of agreement *3390/B/T02/2011/40*.

References

Aldrich C., 1998. *Visualization of transformed multivariate data sets with autoassociative neural networks*. Pattern Recognition Letters, vol. 19, issue: 8, p. 749-764, June.

Asimov D., 1985. *The Grand Tour: A Tool for Viewing Multidimensional Data*. SIAM Journal of Scientific and Statistical Computing, vol. 6, no. 1, p. 128-143.

Assa J., Cohen-Or D., Milo T., 1997. *Displaying data in multidimensional relevance space with 2D visualization maps*. Proceedings. Visualization '97, p. 127-134. New York, NY, IEEE.

Assa J., Cohen-Or D., Milo T., 1999. *RMAP: a system for visualizing data in multidimensional relevance space*. Visual Computer, vol. 15, no. 5, p. 217-234.

Becker R.A., Cleveland W.S., Wilks A.R., 1987. *Dynamic graphics for data analysis*. Statistical Science 2, p. 355-395.

Brożek M., Surowiak A., 2010. *Argument of separation at upgrading in the JIG*. Arch. Min. Sci., Vol. 55, iss. 1, p. 39-40.

Buja A., Asimov D., 1985. *Grand Tour Methods: An Outline*. Computing Science and Statistics, vol. 17, p. 63-67.

Chatterjee A., Das P.P., Bhattacharya S., 1993. *Visualization in linear programming using parallel coordinates*. Pattern Recognition 26(11), p.1725-1736.

Chou S.Y., Lin S.W., Yeh C.S., 1999. *Cluster identification with parallel coordinates*. Pattern Recognition Letters 20, p. 565-572.

Cleveland W.S., McGill R., 1984. *The many faces of a scatterplot*. Journal of the American Statistical Association 79, p. 807-822.

Cook D., Buja A., Cabrera J., Hurley C., 1995. *Grand Tour and Projection Pursuit*. Journal of Computational and Graphical Statistics, vol. 4, no. 3 p. 155-172.

Eick S.G., Wills G.J., 1995. *High interaction graphics*. European Journal of Operational Research, vol. 81, issue: 3, p. 445-459, March 16.

Gennings C., Dawson K.S., Carter W.H., Jr. Myers R.H., 1990. *Interpreting plots of a multidimensional dose-response surface in a parallel coordinate system*. Biometrics 46, p. 719-735.

Hartigan J.A., Kleiner B., 1981. *Mosaic for Contingency Tables*. In: Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface, p. 268-273, New York: Springer Verlag.

Heike H., 2000. *Exploring categorical data: interactive mosaic plots*. Metrika 51, p. 11-26.

Hurley C., Buja A., 1990. *Analyzing high-dimensional data with motion graphic*s. SIAM Journal on Scientific & Statistical Computing, vol. 11, no. 6, p. 1193-1211, Nov.

Inselberg A., Dimsdale B., 1994. *Multidimensional lines I: representation*. SIAM J. Appl. Math. 54 (2), p. 559-577.

Inselberg A., 1985. *The plane with parallel coordinates*, Visual Computer 1, p. 69-91

Jain A.K., Mao J., 1992. *Artificial neural network for non-linear projection of multivariate data*. In: Proc. IEEE Internat. Joint Conf. On Neural Networks, Baltimore, MD, 3, p. 335-340.

Jamróz D., 2009. *Multidimensional labyrinth – multidimensional virtual reality*. In: K. Cyran, S. Kozielski, J. Peters, Stanczyk U., Wakulicz-Deja A. (eds.) Man-Machine, Interactions, AISC, vol. 59, p. 445-450. Springer-Verlag, Berlin Heidelberg, Germany.

Jamróz D., 2001. *Visualization of objects in multidimensional spaces*. Ph.D. Thesis, AGH, University of Science and Technology, Cracow, Poland.

Kim S., Kwon S., Cook D., 2000. *Interactive visualization of hierarchical clusters using MDS and MST*. Metrika 51, p. 39-51, Springer-Verlag.

Kraaijveld M., Mao J., Jain A.K., 1995. *A nonlinear projection method based on Kohonen's topology preserving maps*. IEEE Trans. Neural Networks 6(3), p. 548-559.

Li W., Yue H.H., Valle-Cervantes S., Qin S.J., 2000. *Recursive PCA for adaptive process monitoring*. Journal of Process Control, vol. 10, issue: 5, p. 471-486, October.

Lyman G.J., 1993. *Application of Line-Length Related Interpolation Methods to Problems in Coal Preparation – III: Two dimensional Washability Data Interpolation*. Coal Preparation, vol. 13, p. 179-195.

Mao J., Jain A.K., 1995. *Artificial neural networks for feature extraction and multivariate data projection*. IEEE Trans. Neural Networks 6(2), p. 296-317.

Niedoba T., 2013. *Statistical analysis of the relationship between particle size and particle density of raw coal*. Physicochemical Problems of Mineral Processing, vol. 49, iss. 1, p. 175-188.

Niedoba T., Surowiak A., 2012. *Type of coal and multidimensional description of its composition with density and ash contents taken into consideration*. in Proceedings of the XXVI International Mineral Processing Congress, vol. 1, p. 3844-3854.

Niedoba T., 2011. *Three-dimensional distribution of grained materials characteristics*. in Proceedings of the XIV Balkan Mineral Processing Congress, Tuzla, Bosnia and Herzegovina, vol. 1, p. 57-59.

Niedoba T., 2009. *Wielowymiarowe rozkłady charakterystyk materiałów uziarnionych przy zastosowaniu nieparametrycznych aproksymacji funkcji gęstości rozkładów brzegowych*. Górnictwo i Geoinżynieria, iss. 4, p. 235-244.

Olejnik T., Surowiak A., Gawenda T., Niedoba T., Tumidajski T., 2010. *Wielowymiarowe charakterystyki węgli jako podstawa do oceny i korekty technologii ich wzbogacania*. Górnictwo i Geoinżynieria, vol. 34, iss. 4/1, p. 207-216.

Sobol M.G., Klein G., 1989. *New graphics as computerized displays for human information processing*. IEEE Trans. Systems Man Cybernet. 19(4), p. 893-898.

Stanisz A., 2007. *Przystępny kurs statystyki w oparciu o program Statistica PL na przykładach z medycyny, tom III: Analizy wielowymiarowe*. Wyd. Statsoft, Kraków.

Tumidajski T., Saramak D., 2009. *Metody i modele statystyki matematycznej w przeróbce surowców mineralnych*. Wydawnictwo AGH, Kraków.

Wegman E.J., 1990. *Hyper-dimensional data analysis using parallel coordinates*. J. Amer. Statist. Assoc. 85 (411), p. 664-675.

Received: 10 April 2013